Complexity of Inference in Graphical Models*

Venkat Chandrasekaran[1], Nathan Srebro[2], and Prahladh Harsha[3]

[1] Laboratory for Information and Decision Systems
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

[2] Toyota Technological Institute – Chicago
Chicago, IL 60637 USA

[3] School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai 400005, INDIA

May 7, 2010

Abstract

Graphical models provide a convenient representation for a broad class of probability distributions. Due to their powerful and sophisticated modeling capabilities, such models have found numerous applications in machine learning and other areas. In this paper we consider the complexity of commonly encountered tasks involving graphical models such as the computation of the mode of a posterior probability distribution (i.e., MAP estimation), and the computation of marginal probabilities or the partition function. It is well-known that such *inference* problems are hard in the worst case, but are tractable for models with bounded treewidth. We ask whether treewidth is the only *structural* criterion of the underlying graph that enables tractable inference. In other words, is there some class of structures with unbounded treewidth in which inference is tractable? Subject to a combinatorial hypothesis due to Robertson, Seymour, and Thomas (1994), we show that low treewidth is indeed the *only* structural restriction that can ensure tractability. More precisely we show that for *every* growing family of graphs indexed by tree-width, there exists a choice of potential functions such that the corresponding inference problem is *intractable*. Thus even for the "best case" graph structures of high treewidth, there is no polynomial-time inference algorithm. Our analysis employs various concepts from complexity theory and graph theory, with graph minors playing a prominent role.

Keywords: Graphical models; Markov random fields; treewidth; graph minor; complexity; inference

1 INTRODUCTION

Graphical models provide a powerful formalism for probabilistic modeling. They refer to probability distributions in which the conditional independence structure is represented by a graph, and

*Portions of this work were done while the first and third authors were at the Toyota Technological Institute in Chicago. A preliminary version of this paper appeared in *Proc. 24th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2008 (Chandrasekaran et al., 2008). Email: venkatc@mit.edu; nati@uchicago.edu; prahladh@tifr.res.in hence they are also known as Markov random fields (MRFs). Due to their sophisticated modeling capabilities, graphical models have found application in numerous areas such as computer vision, error-correcting codes, statistical physics, image processing, networking, game theory, and combinatorial optimization. In many of these problems involving graphical models two *inference* tasks are commonly encountered: *(i)* computing the mode or maximum-a-posteriori (MAP) assignment of a posterior probability distribution, and *(ii)* computing the node marginal probabilities or the partition function of a joint probability distribution. In this paper, we study the complexity of these inference problems in graphical models with respect to *structural* properties of the underlying graphs. Specifically, we consider the complexity of inference in a graphical model with discrete-valued random variables as a function of the *treewidth* of the underlying graph.

The notion of treewidth was originally introduced by Robertson and Seymour in their series of papers on graph minors, and it has played a prominent role in a number of results in graph theory (Robertson and Seymour, 1983, 1986; Robertson et al., 1994). The treewidth of a triangulated or chordal graph is one less than the size of the largest clique; the treewidth of a general non-chordal graph is the minimum over the treewidths of all triangulations (see Section 2.3 for formal definitions). Several subsequent papers considered the complexity of hard combinatorial problems on graphs, and showed that many of these problems are tractable in graphs with *bounded* treewidth (Dechter and Pearl, 1989; Freuder, 1990). These ideas extend directly to the graphical model setting – indeed, it is well-known that inference in a graphical model is intractable in the worst-case (Cooper, 1990; Roth, 1996), but is tractable in models with bounded treewidth (Cowell et al., 1999).

A number of methods have been proposed for exact and approximate inference in graphical models; see (Wainwright and Jordan, 2008) and the references therein for the large and growing body of work on this subject. With increasing interest in this problem and in providing conditions under which various procedures are correct and tractable, it is important to understand whether there is indeed some structural property, other than treewidth, which can guarantee tractable inference. For example one recent proposal for exact inference in graphical models is based on an enumeration over so-called "generalized loops" in a graph (Chertkov and Chernyak, 2006), suggesting that inference is tractable in models with a small number of such loops. Another procedure develops conditions based on transformations resulting in perfect graphs under which MAP estimation in a graphical model is tractable (Jebara, 2009). In this paper we consider whether there might be an alternate structural property of graphs, which does not imply low treewidth, but which guarantees tractable inference. In other words, we ask if inference remains hard even in the "easiest" high-treewidth graph structures? We focus purely on structural properties, and consider algorithms that operate with any choice of the potentials.

It is easily seen that inference is hard in models in which the underlying graph is triangulated and has large treewidth, as such a graph contains a large clique as a subgraph. However for models with large treewidth that are not triangulated, it is not apparent whether inference must remain hard. Indeed there are sparse graphs such as the grid that have very large treewidth, but where many edges need to be added to triangulate the graph. In such graphs one might imagine the possibility of efficient procedures that directly take advantage of structure in the graph, rather than first triangulating the graph and then employing the junction-tree algorithm or other similar methods (Cowell et al., 1999).

Recently, Marx showed that constraint satisfaction problems (CSP) defined on *any* class of graphs with unbounded treewidth cannot be solved in polynomial-time unless the exponential time hypothesis[1] fails (Marx, 2007a). This in turn implies that exact inference is hard in every family

[1]The exponential time hypothesis (Impagliazzo et al., 2001) is a commonly believed assumption from complexity

of graphs with unbounded treewidth. However, it is to be noted that Marx's result refers only to algorithms for CSPs involving variables of *unbounded cardinality* (i.e., an unbounded number of states), and hence to inference problems in graphical models in which the cardinality of the variables grows in an unbounded manner with the size of the graphs. Thus, Marx's result is arguably of limited interest for typical inference problems that involve variables with low cardinality or even binary states. Indeed we usually think of the complexity of inference, or even of representation of a graphical model with discrete-valued variables, as growing exponentially with the number of states.

We focus on the complexity of inference in models consisting of *binary* or *ternary* variables defined on any class of graphs with unbounded treewidth. Our main result in this paper is that a hardness result can be obtained even for these models if in addition to the standard complexity assumptions, we assume the well-known *grid-minor hypothesis* from graph minor theory. A minor of a graph $\mathcal{G}$ is a graph $\mathcal{H}$ that can be obtained from $\mathcal{G}$ by a sequence of vertex/edge deletions and/or edge contractions (see Section 2.4 for a precise definition). In a series of over twenty papers, Robertson and Seymour shed light on various aspects of graph minors and proved important results in graph theory. The theorem of greatest relevance to this paper is one that relates graph minors and treewidth: for each $g \times g$ grid-structured graph $\mathcal{G}$, there exists a finite $\kappa_{\mathrm{GM}}(g)$ such that $\mathcal{G}$ is a minor of *all* graphs with treewidth greater than $\kappa_{\mathrm{GM}}(g)$. The best known lower-bound and upper-bound for $\kappa_{\mathrm{GM}}(g)$ are $\Omega(g^2 \log g)$ and $2^{O(g^5)}$ respectively. The *grid-minor hypothesis* states that $\kappa_{\mathrm{GM}}(g)$ is polynomially bounded with respect to g. The hypothesis is based on the belief that $\kappa_{\mathrm{GM}}(g)$ is closer to $\Omega(g^2 \log g)$ than $2^{O(g^5)}$ (Robertson et al., 1994); further evidence in support of this hypothesis is provided by Demaine et al. (2009), and by Reed and Wood (2008).

Main results: We show that it is intractable to compute (even *approximately*) the MAP assignment and the partition function in unbounded treewidth graphical models with discrete-valued random variables:

1. **MAP estimation:** We prove that there is no algorithm to compute the MAP assignment with runtime polynomial in treewidth assuming the grid-minor hypothesis and $\mathbf{NP} \not\subseteq \mathbf{P/poly}$.[2] Furthermore, if we replace the $\mathbf{NP} \not\subseteq \mathbf{P/poly}$ assumption by the stronger exponential-time hypothesis (Impagliazzo et al., 2001), we can show that there is no $(1+\varepsilon)$-approximation scheme for the MAP estimation problem with runtime $2^{\mathcal{O}((\frac{1}{\varepsilon})^{1-\delta})}\mathrm{poly}(\mathrm{tw})$ in a graphical model with treewidth tw for any $\delta > 0$. Here, a $(1+\varepsilon)$-approximation scheme refers to a method to compute an assignment whose log-posterior value approximates the log-posterior value of the MAP assignment[3] to within a multiplicative factor of $1 \pm \varepsilon$. These results hold for graphical models with pairwise interactions and binary-valued random variables.

2. **Partition function:** The hardness of MAP estimation directly implies that the exact computation of the partition function is intractable in graphical models with unbounded treewidth. We further prove that unless either the $\mathbf{NP} \subseteq \mathbf{P/poly}$ assumption or the grid-minor hypothesis fails, there is no fully-polynomial randomized approximation scheme (FPRAS) to compute the partition function in unbounded treewidth graphical models with pairwise interactions and with ternary variables. That is, there is no probabilistic procedure to compute

theory, which states that there exists no algorithm that solves arbitrary instance of n-variable 3-SAT in time $2^{o(n)}$.

[2]The assumption $\mathbf{NP} \not\subseteq \mathbf{P/poly}$ is the non-uniform version of the more popular $\mathbf{NP} \neq \mathbf{P}$ assumption. For more details on uniform vs. non-uniform algorithms, see Section 2.2.1.

[3]Note that obtaining estimates of the log of the maximum posterior value is easier than obtaining estimates of the maximum posterior value; our result states that it is intractable to approximately solve even this easier problem.

the partition function to within a multiplicative factor of $1 \pm \varepsilon$ with success probability $1 - \delta$ that has runtime polynomial in treewidth, in ε^{-1}, *and* in $\log(\frac{1}{\delta})$.

Note that our results demonstrate that the MAP assignment and partition function are hard to compute not only in the *exactly*, but also *approximately*.

For the case of planar graphs, all these results hold without requiring the grid-minor hypothesis assumption. Our results imply that the treewidth of a graphical model is the key *structural* parameter that governs the tractability of inference – inference is tractable in all classes of models with bounded treewidth, while inference is intractable in *every* family of models with unbounded treewidth. We also consider the complexity of inference in graphical models with respect to some arbitrary graph parameter $\alpha(\mathcal{G})$; since treewidth is the critical parameter that governs the complexity of inference, one can answer such questions based on the relationship between the parameter $\alpha(\mathcal{G})$ and treewidth. We elaborate more precisely on this point later in the paper.

Organization: The rest of this paper is organized as follows. Section 2 provides a brief background on inference in graphical models, treewidth, and graph minors. Section 3 presents the formal statement of the problem addressed in this paper. Section 4 describes some combinatorial optimization problems; we prove a reduction from such problems to inference in graphical models, which plays a key role in our analysis. Section 5 provides the main results of this paper. We conclude with a brief discussion in Section 6, and describe some open questions in Section 7.

Relation to previous version of paper: A subset of the results in this paper was presented by Chandrasekaran et al. (2008), who primarily addressed the hardness of *exactly* computing the partition function. A weak inapproximability result was also presented in (Chandrasekaran et al., 2008), which proved that there exists no randomized polynomial-time scheme to approximate the partition function to within an additive constant. In this paper, we expand upon these results to address the tractability of both MAP estimation and computation of the partition function. We also prove stronger hardness-of-approximation results for these problems that guarantee inapproximability to within multiplicative factors.

2 BACKGROUND

2.1 GRAPHICAL MODELS AND INFERENCE

A *graph* $\mathcal{G} = (V, \mathcal{E})$ consists of a set of vertices V and associated edges $\mathcal{E} \subset \binom{V}{2}$, where $\binom{V}{2}$ is the set of all unordered pairs of vertices. A *graphical model* (Lauritzen, 1996) is a collection of random variables indexed by the vertices of a graph; each vertex $v \in V$ corresponds to a random variable x_v, and where for any $A \subset V$, $x_A = \{x_v | v \in A\}$. We assume that each of the variables x_v is discrete-valued with cardinality q. Of interest in this paper are distributions that factor according to a graph $\mathcal{G} = (V, \mathcal{E})$ as follows:

$$p(x_V) = \frac{1}{Z(\psi)} \prod_{v \in V} \psi_v(x_v) \prod_{E \in \mathcal{E}} \psi_E(x_E). \tag{2.1}$$

Here, each ψ_E (or ψ_v) is only a function of the variables x_E (or variable x_v). The functions ψ_v and ψ_E are non-negative and are also known as *potential* or *compatibility* functions. We denote the collection of these potentials by $\psi = \{\psi_v, v \in V\} \cup \{\psi_E, E \in \mathcal{E}\}$. The function $Z(\psi)$ is called the *partition* function and serves to normalize the distribution:

$$Z(\psi) = \sum_{x_V \in \{0,\cdots,q-1\}^{|V|}} \prod_{v \in V} \psi_v(x_v) \prod_{E \in \mathcal{E}} \psi_E(x_E). \tag{2.2}$$

The mode of a distribution is given by

$$\begin{aligned} \widehat{x}^{\mathrm{MAP}} &= \operatorname*{arg\text{-}max}_{x_V \in \{0,\cdots,q-1\}^{|V|}} p(x_V) \\ &= \operatorname*{arg\text{-}max}_{x_V \in \{0,\cdots,q-1\}^{|V|}} \prod_{v \in V} \psi_v(x_v) \prod_{E \in \mathcal{E}} \psi_E(x_E), \end{aligned} \tag{2.3}$$

and we denote the optimal value achieved by $\widehat{x}^{\mathrm{MAP}}$ by

$$\begin{aligned} \psi^{\mathrm{MAP}} &= \max_{x_V \in \{0,\cdots,q-1\}^{|V|}} \prod_{v \in V} \psi_v(x_v) \prod_{E \in \mathcal{E}} \psi_E(x_E) \\ &= \prod_{v \in V} \psi_v(\widehat{x}_v^{\mathrm{MAP}}) \prod_{E \in \mathcal{E}} \psi_E(\widehat{x}_E^{\mathrm{MAP}}). \end{aligned} \tag{2.4}$$

Given a posterior distribution composed of potential functions that factors according to a graph as described above, the following two *inference* tasks are of interest in many applications:

1. **MAP estimation:** The goal here is to compute a configuration of the variables that maximizes the posterior distribution, i.e., to compute $\widehat{x}^{\mathrm{MAP}}$.

2. **Partition function computation:** Another goal is to compute the marginal distribution at some vertex. It is well-known that the complexity of computing the marginal distribution at an arbitrary vertex is comparable to that of computing the partition function. A polynomial-time procedure to solve one of these problems can be used to construct a polynomial-time algorithm for the other. Thus, we consider the complexity of computing the partition $Z(\psi)$.

The intractability of inference arises due to the fact that there are exponentially many terms in the optimization (2.4) and in the sum (2.2). We study the complexity of these inference tasks as a function of structural properties of the underlying graph.

2.2 COMPLEXITY THEORY PRELIMINARIES

We briefly outline some concepts from complexity theory that play an important role in our analysis.

2.2.1 UNIFORM VS. NON-UNIFORM ALGORITHMS

The classical notion of algorithms refers to "uniform algorithms" in which one has a single algorithm that works for all input lengths. A "non-uniform algorithm" on the other hand refers to a family of algorithms, one for each input length. An alternative view of such non-uniform algorithms is that the algorithm is allowed to receive some arbitrary advice that depends only on the input length (but not on the actual input). In the theory of computation literature such "non-uniform algorithms" are usually referred to as fixed-input-size "circuits", where for each input length a different circuit is used (Karp and Lipton, 1980). The class **P** is the class of problems that have polynomial time uniform algorithms while **P/poly** is its non-uniform counterpart, i.e., the class of problems that have polynomial time non-uniform algorithms (circuits). Clearly, $\mathbf{P} \subset \mathbf{P/poly}$. The non-uniform version of the assumption $\mathbf{NP} \neq \mathbf{P}$ is $\mathbf{NP} \not\subseteq \mathbf{P/poly}$, and (though slightly weaker) is equally believed to be true. We need to work with the latter assumption since our proof proceeds by reducing the inference problem on the easiest high-treewidth graphs to a non-uniform algorithm for **NP** (see Section 5 for more details).

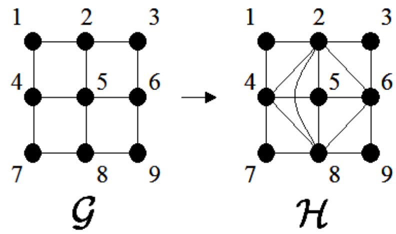

Figure 1: A non-triangulated graph $\mathcal{G}$, and a triangulated supergraph $\mathcal{H}$ of $\mathcal{G}$.

2.2.2 EXPONENTIAL-TIME HYPOTHESIS

Typical hardness results stated with respect to assumptions such as $\mathbf{NP} \neq \mathbf{P}$ lead to conclusions that certain problems do not have polynomial-time algorithms, i.e., every algorithm must have a super-polynomial runtime. To obtain sharper results, we will use the non-uniform version of the so-called "Exponential-Time Hypothesis" (Impagliazzo et al., 2001):

Exponential-time hypothesis (ETH): There exists no non-uniform algorithm[4] that can solve arbitrary instances of n-variable 3-SAT in time $2^{o(n)}$.

The SAT problem will be discussed in greater detail in Section 4. Note that $\mathbf{NP} \not\subseteq \mathbf{P/poly}$ would merely state that there exists no polynomial-time (non-uniform) algorithm for arbitrary n-variable instances of 3-SAT (since 3-SAT is $\mathbf{NP}$-complete). Thus, the ETH is a stronger assumption than $\mathbf{NP} \not\subseteq \mathbf{P/poly}$, and consequently, allows us to obtain sharper bounds on the runtime of inference algorithms (see Section 5 for more details).

2.3 GRAPH TREEWIDTH

A graph is said to be *triangulated* if every cycle of length greater than three contains an edge between two non-adjacent vertices. The *treewidth* $\mathrm{tw}(\mathcal{G})$ of a triangulated graph $\mathcal{G}$ is one less than the size of the largest clique. The treewidth of a general graph is defined

$$\mathrm{tw}(\mathcal{G}) = \min_{\mathcal{H} \supseteq \mathcal{G}, \mathcal{H} \text{ triangulated}} \mathrm{tw}(\mathcal{H}).$$

Here, $\mathcal{H} \supseteq \mathcal{G}$ denotes that $\mathcal{H}$ is a supergraph of $\mathcal{G}$. In words, the treewidth of a graph $\mathcal{G}$ is the minimum over the treewidths of all triangulated supergraphs of $\mathcal{G}$.

Figure 1 shows an example of a non-triangulated graph $\mathcal{G}$, which has a 4-cycle with no edge connecting non-adjacent vertices. The graph $\mathcal{H}$ is a triangulated supergraph of $\mathcal{G}$, and has treewidth 3 as the largest clique is $\{2, 5, 6, 8\}$. Thus, the treewidth of $\mathcal{G}$ is also 3.

The complexity of a graphical model is often measured by the treewidth of the underlying graph, and efficient algorithms are typically known if the underlying graph has low (or constant) treewidth. For instance distributions defined on trees, which have a treewidth equal to one, permit very efficient linear-time inference algorithms. For loopy graphs that have low treewidth the junction-tree method (Cowell et al., 1999) provides an efficient inference algorithm. However, for general loopy graphs the junction-tree method might be intractable because its runtime scales exponentially with the treewidth. As a result considerable effort is being devoted to the development of inference algorithms, and our focus here is on analyzing the computational complexity of both exact and approximate inference.

[4] Impagliazzo et al. (2001) refer to the uniform version of ETH, but their results equally apply to the above-stated non-uniform version of the hypothesis, which is also widely believed to be true.

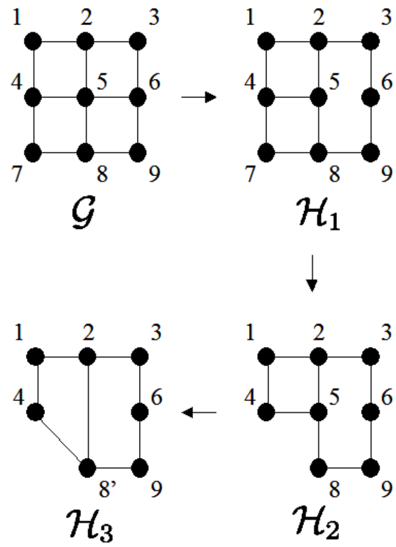

Figure 2: A graph $\mathcal{G}$, and three of its minors $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$ obtained by edge deletion, followed by vertex deletion, and finally edge contraction.

2.4 GRAPH MINORS

The theory of graph minors plays a key role in our analysis. Specifically, we show in Section 5.1 that the complexity of inference in a minor of $\mathcal{G}$ is bounded by the complexity of inference in $\mathcal{G}$. A *minor* of a graph is obtained by any sequence of the following operations:

- **Vertex deletion**: Given a graph $(V, \mathcal{E})$, a vertex $v \in V$ is *deleted*, as are all the edges $\mathcal{E}_v = \{E \in \mathcal{E} : v \in E\}$ incident on v, to obtain the graph $(V \backslash v, \mathcal{E} \backslash \mathcal{E}_v)$.
- **Edge deletion**: Given a graph $(V, \mathcal{E})$, an edge $E \in \mathcal{E}$ is *deleted* to obtain the graph $(V, \mathcal{E} \backslash E)$.
- **Edge contraction**: Given a graph $(V, \mathcal{E})$, an edge $\{u, v\} \in \mathcal{E}$ is *contracted* to form a single vertex u' with edges to every vertex in $V \backslash \{u, v\}$ that previously had an edge to either u or v. In particular the new vertex set becomes $V' = [V \backslash \{u, v\}] \cup \{u'\}$ and the new edge set becomes $\mathcal{E}' = [\mathcal{E} \backslash \{\mathcal{E}_u \cup \mathcal{E}_v\}] \cup \{\{u', w\} \mid w$ previously neighbor of u or $v\}$ Thus, the resulting graph has one less vertex than the original graph.

Figure 2 gives an example of each of these operations. The graph $\mathcal{H}_1$ is a minor of $\mathcal{G}$, and is obtained from $\mathcal{G}$ by deleting the edge $\{5, 6\}$. Next, $\mathcal{H}_2$ is obtained from $\mathcal{H}_1$ by deleting the vertex 7, and the corresponding edges $\{4, 7\}, \{7, 8\}$ that are incident on 7. Thus, $\mathcal{H}_2$ is a minor of both $\mathcal{H}_1$ and $\mathcal{G}$. Finally, $\mathcal{H}_3$ is obtained from $\mathcal{H}_2$ by contracting the edge $\{5, 8\}$ to form the new vertex $8'$, which now has an edge to vertices 2, 4, and 9. The graph $\mathcal{H}_3$ is a minor of each of the graphs $\mathcal{G}$, $\mathcal{H}_1$, and $\mathcal{H}_2$.

In a series of over twenty papers, Robertson and Seymour investigated various aspects of graph minors and proved several important results in graph theory. The following theorem played a key role in proving many of these results; it provides a connection between treewidth and graph minors, and forms a critical component of our analysis.

Theorem 2.1 (Robertson et al. (1994)) *Let $\mathcal{G}$ be a $g \times g$ grid. There exists a* finite $\kappa_{\mathrm{GM}}(g)$ *such that $\mathcal{G}$ is a minor of* all *graphs with treewidth greater than $\kappa_{\mathrm{GM}}(g)$. Further, the best known bounds on $\kappa_{\mathrm{GM}}(g)$ are $c_1 g^2 \log g \leq \kappa_{\mathrm{GM}}(g) \leq 2^{c_2 g^5}$, where c_1 and c_2 are universal constants (i.e., they are independent of g).*

Thus, each grid-structured graph is a minor of all graphs with sufficiently large treewidth. Robertson et al. (1994) expressed the belief that $\kappa_{\mathrm{GM}}(g)$ is closer to $c_1 g^2 \log g$ than $2^{c_2 g^5}$, and may even be on the order of $g^2 \log g$. In addition, several recent results build further support for this conjecture. For example, it is conjectured by Demaine et al. (2009) that $\kappa_{\mathrm{GM}}(g) \sim g^3$. Recently, it was also shown that a graph with polynomial treewidth has a large "grid-like" minor (Reed and Wood, 2008). Consequently, we have the following grid-minor hypothesis.

Grid-minor hypothesis: $\kappa_{\mathrm{GM}}(g)$, as defined in Theorem 2.1, is a polynomial function of g.

This hypothesis is a key assumption in the proofs of our 'Main Results' as stated in the introduction.

Next, we state a restricted result that relates graph minors and treewidth for planar graphs. A planar graph (Bollobás, 1998) is one that can be drawn on a plane with no two edges intersecting each other.

Theorem 2.2 (Robertson et al. (1994)) *There exist universal constants c_3 and c_4 such that the following holds. Let $\mathcal{G}$ be a $g \times g$ grid. Then, (a) $\mathcal{G}$ is a minor of all planar graphs with treewidth greater than $c_3 g$. Further, (b) all planar graphs of size (number of vertices) less than $c_4 g$ are minors of $\mathcal{G}$.*

Hence, Theorem 2.2 states that $\kappa_{\mathrm{GM}}(g)$ is actually *linear* in g for planar graphs.

3 PROBLEM STATEMENT

Let $\mathcal{M}(\mathcal{G}, q)$ refer to the set of all possible choices for potential functions on the vertices and edges of $\mathcal{G} = (V, \mathcal{E})$, with the variables having maximum cardinality q. That is, each $\psi \in \mathcal{M}(\mathcal{G}, q)$ is specified as $\psi = \{\psi_v, v \in V\} \cup \{\psi_E, E \in \mathcal{E}\}$. Let $T_f(I)$ denote the runtime of an algorithm f on input I. We consider exact and approximate inference algorithms that take as input a graph $\mathcal{G} = (V, \mathcal{E})$ and an element of $\mathcal{M}(\mathcal{G}, q)$, and try to compute either the partition function $Z(\psi)$ (2.2) or the MAP estimate $\widehat{x}^{\mathrm{MAP}}$ (2.3) (or the maximum value ψ^{MAP} (2.4)). We would like to investigate the impact of the treewidth $\mathrm{tw}(\mathcal{G})$ of the graph $\mathcal{G}$ on the required runtime of any inference algorithm.

Typical complexity analysis studies the worst case, or maximum, runtime of an algorithm over all inputs. Since inference in a graphical model is **NP**-hard, and assuming **NP** $\neq$ **P**, we know that the worst case runtime of any inference algorithm must scale super-polynomially with the size of the graph. That is, the *maximum* runtime over all graphs is super-polynomial. One can conclude from this that inference is intractable with respect to treewidth in the *worst case*, as one could simply consider the sequence of complete graphs.

Our focus in this paper is on studying the "best case" complexity of inference. We ask whether there is any sequence of graphs $\{\mathcal{G}_k\}$ indexed by treewidth k in which inference runtime scales polynomially with respect to k. We formalize this question as follows. Given a graph $\mathcal{G}$, let

$$\beta_f(\mathcal{G}, q) = \max_{\psi \in \mathcal{M}(\mathcal{G}, q)} T_f(\mathcal{G}, \psi) \tag{3.1}$$

denote the complexity of inference in models defined with respect to $\mathcal{G}$. Given a family of graphs $\{\mathcal{G}_k\}$ indexed by treewidth k, one can consider the quantity $\beta_f(\mathcal{G}_k, q)$ and ask how it scales with k.

Main Question: Is there a family of graphs $\{\mathcal{G}_k\}_{k=1}^{\infty}$ indexed by treewidth k and an inference algorithm f for which $\beta_f(\mathcal{G}_k, q)$ scales polynomially with respect to k?

Since we are primarily concerned with bounds that are independent of the cardinality q, we will specifically consider the cases $q = 2, 3$. If there exists an f such that $\beta_f(\mathcal{G}_k, q)$ is polynomial in k, then there exists a class of structures with unbounded treewidth in which inference would be tractable. Alternatively, if $\beta_f(\mathcal{G}_k, q)$ is not polynomial in k for any procedure f and for any family of graphs $\{\mathcal{G}_k\}$, then bounding the treewidth is the *only* structural restriction on graphical models that leads to tractable inference.

The quantity β_f in the 'Main Question' refers to a uniform algorithm, i.e., a single algorithm that should work for graphs of all treewidths. However, to answer this question we will actually study a slightly harder question, where we allow non-uniform algorithms specialized to a sequence of graphs of increasing treewidths. Given a sequence of graphs $\{\mathcal{G}_k\}_{k=1}^{\infty}$ with $\mathrm{tw}(\mathcal{G}_k) = k$, we will analyze the runtime of any (non-uniform) sequence $f = \{f_k\}_{k=1}^{\infty}$ of algorithms (i.e. a "non-uniform algorithm"), where f_k solves the inference problem on $\mathcal{G}_k$. For any such sequence, we study how the runtime increases (taking worst case over potential functions) with k, i.e., $\max_{\psi \in \mathcal{M}(\mathcal{G}_k)} T_{f_k}(\mathcal{G}_k, \psi)$ as a function of k.

Our 'Main Question' pertains to both exact and approximate inference. For the approximate inference problem, we consider inference algorithms that provide a $(1+\varepsilon)$-approximation either to the log-optimal value $\log \psi^{\mathrm{MAP}}$ or to the partition function $Z(\psi)$. A $(1+\varepsilon)$-approximation procedure for MAP estimation provides an assignment with log-posterior value $\log \widehat{\psi}$ such that $(1-\varepsilon)\log \psi^{\mathrm{MAP}} \le \log \widehat{\psi} \le (1+\varepsilon)\log \psi^{\mathrm{MAP}}$. Similarly a $(1+\varepsilon)$-approximation algorithm for the partition function provides an approximate result $\widehat{Z}$ to within a multiplicative factor of $1 \pm \varepsilon$ of $Z(\psi)$, i.e., $(1-\varepsilon)Z(\psi) \le \widehat{Z} \le (1+\varepsilon)Z(\psi)$; such an approximation algorithm for the partition function can equivalently be viewed as providing additive-factor approximations to the log-partition function. Inference methods that scale polynomially with both the treewidth and the inverse of the approximation factor ε^{-1} are called *fully-polynomial time approximation schemes* (FPTAS). We also study randomized approximation schemes parametrized by an additional parameter $\delta > 0$ that provide the above approximation guarantees with a success probability greater than $1-\delta$. Randomized methods in which the runtime scales polynomially with treewidth, the inverse of the approximation factor ε^{-1}, *and* $\log(\frac{1}{\delta})$ are called *fully-polynomial time randomized approximation schemes* (FPRAS).

The additional parameters ε and δ can be naturally incorporated into the definition of β_f above for a procedure f that provides an approximate result. Thus, the approximation version of our main question asks whether there exists a sequence of graphs $\{\mathcal{G}_k\}$ and a (non-uniform) approximation algorithm $f = \{f_k\}$ for which $\beta_f(\mathcal{G}_k, q, \varepsilon)$ (or $\beta_f(\mathcal{G}_k, q, \varepsilon, \delta)$) scales polynomially with k and ε^{-1} (or with k, ε^{-1}, and $\log(\frac{1}{\delta})$).

Remark: As inference in arbitrary graphical models is **NP**-hard, one can construct a sequence of graphs $\{\mathcal{G}_k\}$ indexed by treewidth in which inference is intractable (assuming $\mathbf{NP} \neq \mathbf{P}$). Indeed as described previously the sequence of complete graphs is one such sequence. Using this reasoning, it is clear that inference is intractable in any sequence of *triangulated* graphs $\{\mathcal{G}_k\}$ parametrized by treewidth. The reason for this is that a triangulated graph with treewidth equal to k must contain a clique of size $k+1$. Therefore, candidates for families of graphs of unbounded treewidth in which inference might be tractable must be far from being triangulated. For example grid-structured graphs provide a candidate family – a $k \times k$ grid has treewidth equal to k, but is far from being triangulated and many edges need to be added in order to triangulate the grid as it is very sparse. Another interesting family is the class of hypercube graphs, where a d-dimensional hypercube consists of 2^d vertices and $2^{d-1}d$ edges, and the treewidth is lower-bounded by $\Omega(\frac{2^d}{d})$ (Chandran and Subramanian, 2003). Such graphs are also far from being triangulated, as the triangulation of a d-dimensional hypercube would consist of $\Omega(\frac{2^{2d}}{d^2})$ edges (while the original graph has just $2^{d-1}d$ edges). For such families of very sparse, structured graphs with large treewidth, one can imagine specialized inference procedures that directly take advantage of special structure in the graphs as opposed to first triangulating and then using standard procedures such as the junction-tree method (Cowell et al., 1999). Our 'Main Question' asks whether there exists any family of graphs, perhaps with some special structure, of unbounded treewidth in which inference is tractable.

4 COMBINATORIAL OPTIMIZATION AND INFERENCE

Combinatorial optimization problems provide a rich source of intractable problems, and proving the hardness of a problem is often done by demonstrating a reduction from one of the hard combinatorial optimization problems. Many combinatorial optimization problems are defined with respect to graphs, and can often be expressed as a *constraint satisfaction problem* (CSP). A CSP is defined as a set of constraints specified on subsets of a collection of discrete-valued variables. Each constraint is said to be *satisfied* for some stipulated configurations of the variables in the constraint. The problem is to identify a configuration of the variables that satisfies all the constraints (i.e., find a *satisfying* assignment). We will mostly be concerned with 2-CSPs: CSPs in which each constraint involves only two variables. Note that one can associate a graph with an instance of a 2-CSP, with the vertices representing the variables and edges present only between those vertices that appear in the same constraint. Of interest to us will be two important variants of a CSP. The first is MAXCSP in which one wants to find an assignment that *maximizes* the number of satisfied constraints. We describe how such problems can be transformed into a MAP estimation problem in a graphical model. The second is #CSP in which one wants to *count* the number of assignments that satisfy all the constraints; such problems are more naturally related to computing the partition function.

An important special case of a CSP is the SAT problem, in which disjunctive constraints are specified on binary variables. Although polynomial time algorithms exist for 2-SAT, the MAX-2-SAT problem is **NP**-complete. In fact we have that planar MAX-2-SAT, in which instances are restricted to those defined on planar graphs, is also **NP**-complete (Guibas et al., 1993).

We will also make use of hardness results for the independent set problem and the colorability problem. An *independent set* of a graph is a subset of the vertices of the graph such that no two vertices are adjacent. A 3-*coloring* of a graph is a mapping from the vertex set to three "colors" $\{R, G, B\}$ such that no two adjacent vertices are assigned the same color. Both these problems can be expressed as 2-CSPs, with a constraint for each edge of the graph specifying either that the "colors" assigned to the two vertices in the edge must not be the same, or that at most one of the vertices in the edge is chosen in the independent set.

4.1 CONSTRAINT SATISFACTION TO INFERENCE

In order to translate hardness results for CSPs and MAXCSPs to the problem of inference in graphical models, we describe transformations from instances of 2-CSPs to inference problems in graphical models. We begin by showing that a MAXCSP problem can be reduced to a MAP estimation problem as follows.

Lemma 4.1 *There exists a polynomial time reduction mapping instances $I = (x_1, \ldots, x_n; \mathcal{R})$ of a MAX-2-CSP problem (here $x_1, \ldots, x_n$ are discrete valued variables of cardinality q and $\mathcal{R}$ is a set of constraints) to a set of potentials $\psi \in \mathcal{M}(\mathcal{G}, q)$ where $\mathcal{G} = (V, \mathcal{E})$ denotes the graph that* *represents the instance I such that the following holds: At least d constraints in $\mathcal{R}$ can be satisfied simultaneously if and only if* $\log \psi^{\mathrm{MAP}} \geq d$.

Proof Let $\mathcal{G} = (V, \mathcal{E})$ denote the graph which represents the instance I. Here $|V| = n$ with each variable being assigned to a vertex and $\mathcal{E}$ contains only those pairs of vertices for which the corresponding variables appear in the same relation, so that $|\mathcal{E}| = |\mathcal{R}|$. For each $E \in \mathcal{R}$, define

$$\psi_E(x_E) = \begin{cases} e, & x_E \text{ satisfies } E \\ 1, & \text{otherwise.} \end{cases}$$

Define vertex potentials similarly for each vertex constraint, and set $\psi_v = 1$ for other vertices. It is clear that $\psi^{\mathrm{MAP}} \geq e^d$ if and only if at least d constraints can be simultaneously satisfied in I. ■

Remark: The maximum independent set problem can be transformed to a MAP estimation problem in a graphical model (with binary variables) using a slightly different construction to the one presented in this lemma. Specifically, given a graph $\mathcal{G} = (V, \mathcal{E})$ consider the following set of potentials for each $v \in V$ and for each $E \in \mathcal{E}$:

$$\begin{aligned} \psi_v(x_v) &= e^{x_v} \\ \psi_E(x_E) &= \begin{cases} 1, & x_E = (1,1) \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Here each variable takes on the values $\{0, 1\}$. It is then easily seen that $\log \psi^{\mathrm{MAP}} \geq d$ if and only if there is an independent set of size at least d in $\mathcal{G}$.

In Section 5.2, we use this lemma along with the hardness of planar maximum independent set to conclude that MAP estimation is intractable in models with unbounded treewidth. One can also transform instances of #CSP to an instance of computing the partition function in a graphical model.

Lemma 4.2 *There exists a polynomial time reduction mapping instances $I = (x_1, \cdots, x_n; \mathcal{R})$ of a #2-CSP problem (here $x_1, \cdots, x_n$ are discrete-valued variables of cardinality q and $\mathcal{R}$ is a set of constraints) to a set of potentials $\psi \in \mathcal{M}(\mathcal{G}, q)$ where $\mathcal{G} = (V, \mathcal{E})$ denotes the graph that represents the instance I such that the following holds: There exist at least d distinct satisfying assignments (i.e., d distinct assignments of the variables that satisfy all the constraints) if and only if $\lfloor Z(\psi) \rfloor \geq d$.*

Proof As with the previous proof, let $\mathcal{G} = (V, \mathcal{E})$ denote the graph which represents the instance I. Hence, $|V| = n$ with each variable being assigned to a vertex and $\mathcal{E}$ contains only those pairs of vertices for which the corresponding variables appear in the same relation, so that $|\mathcal{E}| = |\mathcal{R}|$. For each $E \in \mathcal{R}$, define

$$\psi_E(x_E) = \begin{cases} 1, & x_E \text{ satisfies } E \\ \alpha, & \text{otherwise.} \end{cases}$$

Define vertex potentials similarly for each vertex constraint, and set $\psi_v = 1$ for other vertices. Here, we choose $\alpha \in \left(0, \frac{1}{q^n}\right)$. If there are at least d satisfying assignments, it is clear that $\lfloor Z(\psi) \rfloor \geq d$. Alternatively if there are fewer than d satisfying assignments, we have that

$$Z(\psi) \quad = \sum_{x_V \in \{0, \cdots, q-1\}^n} \psi(x) \quad \leq \quad (d-1) + \sum_{x_V \in \{0, \cdots, q-1\}^n} \alpha \quad \leq d - 1 + q^n \alpha < d.$$

Consequently, we have that $\lfloor Z(\psi) \rfloor < d$ as d is an integer. ■

In Section 5.3, we use this lemma along with the hardness of counting the number of 3-colorings in planar graph to conclude the hardness of computing the partition function in models with unbounded treewidth. By setting d equal to the total number of constraints in $\mathcal{R}$ in Lemma 4.1, one can transform a decision version of the CSP problem to a particular decision version of the MAP estimation problem. By setting $d = 1$ in Lemma 4.2, one can do the same with the partition function computation.

4.2 HARDNESS OF INFERENCE IN UNBOUNDED CARDINALITY MODELS

Next, we translate recent hardness results by Marx (2007a) for 2-CSPs to a complexity result for inference.

Theorem 4.3 (Marx (2007a)) [5] *Let $\{\mathcal{G}_k\}_{k=1}^{\infty}$ be* any *sequence of graphs indexed by treewidth. Suppose that there exists an algorithm g for instances of 2-CSPs, with variables of arbitrary cardinality, defined on the graphs $\mathcal{G}_k$. Let $q(\psi)$ be the maximum cardinality of a variable referred to by the constraints ψ. If $T_g(\mathcal{G}_k, \psi) = q(\psi)^{o(\frac{k}{\log k})}$, then the ETH fails.*

Corollary 4.4 *Let f be any algorithm that can perform inference on graphical models defined on a family of graphs $\{\mathcal{G}_k\}_{k=1}^{\infty}$ indexed by treewidth with variables of arbitrary cardinality. Under the ETH, for any $r(k) = o(k/\log k)$ there exist q, k such that $\beta_f(\mathcal{G}_k, q) > q^{r(k)}$.*

Proof From Lemma 4.1, we have that arbitrary instances of 2-CSP can be transformed to a decision version of an inference problem in a graphical model in polynomial-time. Consequently, we have that if there exists an inference algorithm that can perform inference in time $q^{o(\frac{k}{\log k})}$, then the ETH fails. ■

A consequence of this corollary is that the junction-tree algorithm (Cowell et al., 1999), which scales as q^k, is in a sense near-optimal (assuming the ETH). However, as we noted in the introduction this result provides an asymptotic lower bound only for sufficiently large cardinalities. It does not provide a lower bound for any fixed cardinality q. This restriction plays an important role in the reductions of Marx (2007a), in which large sets of variables in an intermediate model are represented using a single high-cardinality variable. In the following section, we describe our main results for the complexity of inference in graphical models with *binary* and *ternary* variables, which are typically of most interest to the machine learning community.

5 MAIN RESULTS

We present our main results for graphical models with binary-valued and ternary-valued variables in this section. We begin by analyzing the complexity of inference in a graphical model defined on graph $\mathcal{G}$ in terms of the complexity of inference in models defined on minors of $\mathcal{G}$. Next, we analyze the complexity of MAP estimation and of partition function computation in families of models with unbounded treewidth. Finally, we also consider the restricted case of inference in planar graphs of unbounded treewidth for which we prove complexity bounds without recourse to the grid-minor hypothesis. We consider both the exact and approximate versions of these tasks.

[5] The statement here is actually the non-uniform variant of the result of Marx (2007a).

5.1 INFERENCE AND GRAPH MINORS

In the following two propositions, we relate the complexity of inference in a minor of a graph $\mathcal{G}$ to inference in $\mathcal{G}$. First, we show that the problem of computing the partition function in a minor of a graph $\mathcal{G}$ can be transformed to a problem in $\mathcal{G}$.

Proposition 5.1 *Let $\mathcal{H}$ be a minor of $\mathcal{G}$, and let $\psi_{\mathcal{H}} \in \mathcal{M}(\mathcal{H}, q)$. There exists a $\psi_{\mathcal{G}} \in \mathcal{M}(\mathcal{G}, q)$ such that $Z(\psi_{\mathcal{H}}) = Z(\psi_{\mathcal{G}})$. Moreover, $\psi_{\mathcal{G}}$ can be computed in linear time given $\psi_{\mathcal{H}}$ and the sequence of minor operations that transform $\mathcal{G}$ to $\mathcal{H}$.*

Proof All we need to show is that if a graph $\mathcal{H} = (V_{\mathcal{H}}, \mathcal{E}_{\mathcal{H}})$ is obtained from another graph $\mathcal{G} = (V_{\mathcal{G}}, \mathcal{E}_{\mathcal{G}})$ by just a *single* application of one of the standard minor operations, then we can transform a given $\psi_{\mathcal{H}} \in \mathcal{M}(\mathcal{H}, q)$ into a $\psi_{\mathcal{G}} \in \mathcal{M}(\mathcal{G}, q)$ with $Z(\psi_{\mathcal{G}}) = Z(\psi_{\mathcal{H}})$.

Vertex deletion: Suppose that $v \in V_{\mathcal{G}}$ as well as edges $\mathcal{E}_v \subseteq \mathcal{E}_{\mathcal{G}}$ that are incident on v in $\mathcal{G}$ are deleted. Let $\psi_v = \frac{1}{q}$ and let $\psi_E = 1, \forall E \in \mathcal{E}_v$. Letting $\psi_{\mathcal{G}} = \cup_{E \in \mathcal{E}_v} \psi_E \cup \psi_v \cup \psi_{\mathcal{H}}$, one can check that $Z(\psi_{\mathcal{G}}) = Z(\psi_{\mathcal{H}})$.

Edge deletion: Suppose that $E \in \mathcal{E}_{\mathcal{G}}$ is deleted. Setting $\psi_E = 1$, and $\psi_{\mathcal{G}} = \psi_{\mathcal{H}} \cup \psi_E$, one can check that $Z(\psi_{\mathcal{G}}) = Z(\psi_{\mathcal{H}})$.

Edge contraction: Suppose that $\{u, v\} \in \mathcal{E}_{\mathcal{G}}$ is contracted to form the new vertex $u' \in V_{\mathcal{H}}$. We define $\psi_{\{u,v\}}(x_u, x_v) = \delta(x_u - x_v)$, where $\delta(\cdot)$ is the Kronecker delta function that evaluates to 1 if the argument is 0, and 0 otherwise. For the edge potentials, if a vertex $w \in V_{\mathcal{G}} \backslash \{u, v\}$ is originally connected in $\mathcal{G}$ by an edge to only one of u or v, then we set the corresponding $\psi_{\{u,w\}}$ or $\psi_{\{v,w\}}$ to be equal to $\psi_{\{u',w\}}$. If *both* u and v are originally connected by edges to w in $\mathcal{G}$, then we define $\psi_{\{u,w\}} = \psi_{\{u',w\}}$ and $\psi_{\{v,w\}} = 1$. Finally, we define the vertex potentials as $\psi_u = \psi_{u'}$ and $\psi_v = 1$. Letting all the other vertex and edge potentials in $\mathcal{G}$ be the same as those in $\mathcal{H}$, it is easily seen that $Z(\psi_{\mathcal{G}}) = Z(\psi_{\mathcal{H}})$. ■

Next, we prove a similar result for the MAP estimation problem.

Proposition 5.2 *Let $\mathcal{H}$ be a minor of $\mathcal{G}$, and let $\psi_{\mathcal{H}} \in \mathcal{M}(\mathcal{H}, q)$. There exists a $\psi_{\mathcal{G}} \in \mathcal{M}(\mathcal{G}, q)$ such that $\psi_{\mathcal{H}}^{\mathrm{MAP}} = \psi_{\mathcal{G}}^{\mathrm{MAP}}$. Moreover, $\psi_{\mathcal{G}}$ can be computed in linear time given $\psi_{\mathcal{H}}$ and the sequence of minor operations that transform $\mathcal{G}$ to $\mathcal{H}$.*

Proof This proof proceeds in a similar manner to the previous proof. We show that if a graph $\mathcal{H} = (V_{\mathcal{H}}, \mathcal{E}_{\mathcal{H}})$ is obtained from another graph $\mathcal{G} = (V_{\mathcal{G}}, \mathcal{E}_{\mathcal{G}})$ by just a *single* application of one of the standard minor operations, then we can transform a given $\psi_{\mathcal{H}} \in \mathcal{M}(\mathcal{H}, q)$ into a $\psi_{\mathcal{G}} \in \mathcal{M}(\mathcal{G}, q)$ with $\psi_{\mathcal{G}}^{\mathrm{MAP}} = \psi_{\mathcal{H}}^{\mathrm{MAP}}$.

Vertex deletion: Suppose that $v \in V_{\mathcal{G}}$ as well as edges $\mathcal{E}_v \subseteq \mathcal{E}_{\mathcal{G}}$ that are incident on v in $\mathcal{G}$ are deleted. Let $\psi_v = 1$ and let $\psi_E = 1, \forall E \in \mathcal{E}_v$. Letting $\psi_{\mathcal{G}} = \cup_{E \in \mathcal{E}_v} \psi_E \cup \psi_v \cup \psi_{\mathcal{H}}$, one can check that $\psi_{\mathcal{G}}^{\mathrm{MAP}} = \psi_{\mathcal{H}}^{\mathrm{MAP}}$.

Edge deletion: Suppose that $E \in \mathcal{E}_{\mathcal{G}}$ is deleted. Setting $\psi_E = 1$, and $\psi_{\mathcal{G}} = \psi_{\mathcal{H}} \cup \psi_E$, one can check that $\psi_{\mathcal{G}}^{\mathrm{MAP}} = \psi_{\mathcal{H}}^{\mathrm{MAP}}$.

Edge contraction: Suppose that $\{u, v\} \in \mathcal{E}_{\mathcal{G}}$ is contracted to form the new vertex $u' \in V_{\mathcal{H}}$. Defining the potentials in the same manner as in the partition function case above, it is easily seen that $\psi_{\mathcal{G}}^{\mathrm{MAP}} = \psi_{\mathcal{H}}^{\mathrm{MAP}}$. ■

Thus, these results allow us to establish hardness of inference in a graph $\mathcal{G}$, by establishing hardness of inference in a minor of $\mathcal{G}$, provided we know the sequence of minor operations that transform $\mathcal{G}$ to its minor.

5.2 MAP ESTIMATION

In this section we consider the hardness of exact and approximate inference in families of graphical models with unbounded treewidth. Our analysis proceeds by using Theorems 2.1 and 2.2 along with the grid-minor hypothesis to show that *all* graphs with sufficiently large treewidth contain *every* planar graph up to a certain size as a minor. Consequently, we conclude that MAP estimation is hard in graphical models with unbounded treewidth by appealing to the intractability of MAP estimation in planar graphs and the grid-minor hypothesis.

Theorem 5.3 *Let $\{\mathcal{G}_k\}_{k=1}^{\infty}$ be an infinite sequence of graphs indexed by treewidth. Let $f = \{f_k\}_{k=1}^{\infty}$ be any (possibly non-uniform) sequence of algorithms that solves the decision version of the MAP estimation problem on $\{\mathcal{G}_k\}$ with binary variables, i.e., deciding whether $\log \psi^{\text{MAP}} \geq c$ for a set of potentials ψ and for* any *c.*

(a) Assuming that $\mathbf{NP} \not\subseteq \mathbf{P/poly}$ *and that the grid-minor hypothesis holds, $\beta_f(\mathcal{G}_k, 2)$ is super-polynomial in k.*

(b) Assuming that $\kappa_{\text{GM}}(g) = O(g^r)$ in the grid-minor hypothesis and the ETH, we have that $\beta_f(\mathcal{G}_k, 2) = 2^{\Omega(k^{1/2r})}$.

Proof (a) Suppose for the sake of a contradiction that there exists a (possibly non-uniform) polynomial time algorithm f that solves the inference problem on $\{\mathcal{G}_k\}_{k=1}^{\infty}$. More precisely, let $f = \{f_k\}_{k=1}^{\infty}$ be a sequence of algorithms such that f_k solves the inference problem on $\mathcal{G}_k$ in polynomial time. Assuming the grid-minor hypothesis, we will demonstrate that this implies a non-uniform polynomial time algorithm for the inference problem on any planar graph. Recall that planar MAX-2-SAT is **NP**-complete (Guibas et al., 1993), and polynomial-time reducible to the inference problem on planar graphs (Lemma 4.1). This provides a (non-uniform) polynomial time algorithm for an **NP**-complete problem, contradicting the $\mathbf{NP} \not\subseteq \mathbf{P/poly}$ assumption.

Given an instance $(\mathcal{G}, \psi)$ of the inference problem on planar graphs, we proceed as follows: Let $|\mathcal{G}| = s$. By Theorem 2.2, $\mathcal{G}$ is a minor of the $s/c_4 \times s/c_4$ grid. Furthermore, the sequence of minor operations that transform a $s/c_4 \times s/c_4$ grid to $\mathcal{G}$ can be obtained in polynomial time (Tamassia and Tollis, 1989). Thus, using Lemma 5.2 the inference problem $(\mathcal{G}, \psi)$ can be reduced to an inference problem on the $s/c_4 \times s/c_4$ grid in time linear in s. By Theorem 2.1, the $s/c_4 \times s/c_4$ grid is a minor of $\mathcal{G}_{\kappa_{\text{GM}}(s/c_4)}$. We will now use as "non-uniform advice" the sequence of minor operations that transform $\mathcal{G}_{\kappa_{\text{GM}}(s/c_4)}$ to the $s/c_4 \times s/c_4$ grid. Note that this depends only on the input size s and not on the actual instance $(\mathcal{G}, \psi)$. Using Lemma 5.2 again, we can reduce the inference problem on the $s/c_4 \times s/c_4$ grid to an inference problem on $\mathcal{G}_{\kappa_{\text{GM}}(s/c_4)}$ in linear time. We now use $f_{\kappa_{\text{GM}}(s/c_4)}$ to solve the inference problem on $\mathcal{G}_{\kappa_{\text{GM}}(s/c_4)}$, thus solving the original inference problem $(\mathcal{G}, \psi)$. The fact that $\beta_f(\mathcal{G}_k, 2)$, and thus also the size of the graph $\mathcal{G}_k$, is at most polynomial in k and the grid-minor hypothesis (i.e, $\kappa_{\text{GM}}(g) = \text{poly}(g)$) imply that the above algorithm is a polynomial time (non-uniform) algorithm for the inference problem on planar graphs.

(b) We obtain the tighter hardness result by carefully analyzing the running time of the inference algorithm on planar graphs suggested in (a). It can be easily checked that the above algorithm runs in time $\beta_f(\mathcal{G}_{\kappa_{\text{GM}}(s/c_4)}, 2)$, which is $\beta_f(\mathcal{G}_{O(s^r)}, 2)$ if $\kappa_{\text{GM}}(g) = O(g^r)$. Combining this with the reduction from 3-SAT to planar MAX-2-SAT (Litchenstein, 1982; Guibas et al., 1993), which blows up the instance size by a quadratic factor, we obtain a $\beta_f(\mathcal{G}_{O(k^{2r})}, 2)$ time non-uniform algorithm for k-variable instances of 3-SAT. Recall that the ETH states there exists no (non-uniform) algorithm for arbitrary k-variable instances of 3-SAT that has running time $2^{o(k)}$. Hence, assuming the grid-minor hypothesis and the ETH, we must have that $\beta_f(\mathcal{G}_{O(k^{2r})}, 2)$ is at least $2^{\Omega(k)}$ or equivalently that $\beta_f(\mathcal{G}_k, 2)$ is at least $2^{\Omega(k^{1/2r})}$. ■

Notice that the ETH assumption enables a sharper performance bound instead of the simpler result that $\beta_f(k, 2)$ is super-polynomial in k (of part (a)). Theorem 5.3 provides an answer to the hardness of *exact* MAP estimation. Next, we consider the hardness of approximate MAP estimation in graphical models.

Theorem 5.4 *Let $\{\mathcal{G}_k\}_{k=1}^{\infty}$ be an infinite sequence of graphs indexed by treewidth. Let $f = \{f_k\}_{k=1}^{\infty}$ be any (possibly non-uniform) sequence of approximate algorithms operating on models defined on $\{\mathcal{G}_k\}$ with binary variables that compute an approximation to the log-optimal value $\log \psi^{\mathrm{MAP}}$ to within a multiplicative factor of $1 \pm \varepsilon$, i.e., compute an assignment with posterior value $\widehat{\psi}$ such that $(1-\varepsilon)\log \psi^{\mathrm{MAP}} \leq \log \widehat{\psi} \leq (1+\varepsilon)\log \psi^{\mathrm{MAP}}$. Assuming the exponential-time hypothesis, the procedure f cannot have runtime $2^{\mathcal{O}((\frac{1}{\varepsilon})^{1-\delta})}\mathrm{poly(k)}$ for any $\delta > 0$. In particular, f cannot be a fully-polynomial time approximation scheme, i.e., $\beta_f(\mathcal{G}_k, 2)$ cannot be polynomial in both k and ε^{-1}.*

Proof This proof proceeds in a similar manner to the proof above. We use the fact that under the exponential-time hypothesis there is no procedure with runtime $2^{\mathcal{O}((\frac{1}{\varepsilon})^{1-\delta})}\mathrm{poly}(n)$ to compute a $(1+\varepsilon)$-approximation to the maximum independent set in planar graphs of size n (Marx, 2007b). See the remark following Lemma 4.1 to transform an instance of the maximum independent set problem to an instance of the MAP estimation problem. ∎

5.3 PARTITION FUNCTION COMPUTATION

The results in the previous section show that even approximate MAP estimation is intractable in any family of graphical models with unbounded treewidth. In this section we study the hardness of computing the partition function both exactly and approximately. Based on the hardness of exact/approximate MAP estimation, it is clear that *exactly* computing the partition function in unbounded treewidth graphical models is also intractable. We state the following result without proof.

Theorem 5.5 *Let $\{\mathcal{G}_k\}_{k=1}^{\infty}$ be an infinite sequence of graphs indexed by treewidth. Let $f = \{f_k\}_{k=1}^{\infty}$ be any (possibly non-uniform) sequence of algorithms that solve a decision version of the partition function computation problem in models defined on $\{\mathcal{G}_k\}$ with binary variables, i.e., deciding whether $Z(\psi) \geq c$ for a set of potentials ψ and* any *c.*

(a) Assuming that **NP** $\not\subseteq$ **P/poly** *and that the grid-minor hypothesis holds, $\beta_f(\mathcal{G}_k, 2)$ is super-polynomial in k.*

(b) Assuming that $\kappa_{\mathrm{GM}}(g) = O(g^r)$ in the grid-minor hypothesis and the ETH, we have that $\beta_f(\mathcal{G}_k, 2) = 2^{\Omega(k^{1/2r})}$.

It is possible to derive the results in this theorem using weaker assumptions. Specifically the first part can be proved by making the weaker assumption that there do not exist non-uniform polynomial time algorithms for solving counting problems in $\#P$. Similarly the second part can be proved by assuming that there is no non-uniform procedure to count the number of satisfying assignments of arbitrary n-variable instances of 3-SAT in time $2^{o(n)}$. Next we consider the hardness of approximately computing the partition function in models with unbounded treewidth. In particular we show that it is intractable to approximate the partition function up to a multiplicative factor in graphical models with unbounded treewidth consisting of *ternary* variables. The proof proceeds by using a reduction from planar 3-colorability, which is known to be NP-complete.

Theorem 5.6 *Let $\{\mathcal{G}_k\}_{k=1}^{\infty}$ be an infinite sequence of graphs indexed by treewidth. Let $f = \{f_k\}_{k=1}^{\infty}$ be any (possibly non-uniform) sequence of randomized algorithms that compute the partition function $Z(\psi)$ to within a multiplicative factor of $1 \pm \varepsilon$ with probability greater than $1-\delta$, i.e., compute a $\widehat{Z}$ such that $(1-\varepsilon)Z(\psi) \leq \widehat{Z} \leq (1+\varepsilon)Z(\psi)$. Assuming that* **NP** $\not\subseteq$ **P/poly** *and that the grid-minor hypothesis holds, the procedure f cannot have runtime polynomial in the treewidth k, ε^{-1},* and *in $\log(\frac{1}{\delta})$, i.e., f cannot be a fully-polynomial time randomized approximation scheme.*

Proof We have that planar 3-colorability is NP-complete (Stockmeyer, 1973). Consequently, there is no fully-polynomial time randomized approximation scheme to *count* the number 3-colorings in a planar graph (Jerrum, 2003). Using Lemma 4.2 to reduce a counting problem to a partition function computation problem, we can prove the statement of this theorem by following a similar proof to that of Theorem 5.3. It is to be noted that following the same proof structure leads to a *randomized* polynomial time non-uniform algorithm for an **NP**-complete problem and not a *deterministic* polynomial time non-uniform algorithm as in the previous case. However, this suffices to complete the proof as the existence of such a randomized algorithm implies that **NP** $\subseteq$ **BPP/poly**[6] which when combined with the fact that **BPP/poly** = **P/poly** contradicts the **NP** $\not\subseteq$ **P/poly** assumption. ∎

Note that unlike all of our previous results, we only show the hardness of (approximate) partition function computation for ternary variables. Although the number of variables is still finite in our result unlike that in Marx's result (which also only deals with exact inference), proving such a result with binary variables is of interest. A proof of a similar result for binary variables could take several forms. One technique might be to show the hardness of approximate counting for some problem in a planar graph that can be transformed easily into a partition function computation problem in a binary (planar) graphical model. Alternatively, one could use some result on the NP-completeness of a *binary* decision problem on planar graphs that can be transformed easily to an inference problem. Note that while problems with binary variables such as planar 3-SAT are NP-complete, they are not defined on *pairwise* graphs, i.e., graphs with pairwise edges that are of most interest in graphical modeling.

5.4 INFERENCE IN PLANAR GRAPHS

Next, we consider the hardness of inference in graphical models defined on planar graphs. The grid-minor hypothesis played a key role in the results in the previous sections. However if we restrict our attention to *planar* graphs, we do not need to appeal to the grid-minor hypothesis due to Theorem 2.2. We have the following theorem on the hardness of exact and approximate inference in models defined on families of planar graphs.

Theorem 5.7 *Let $\{\mathcal{G}_k\}_{k=1}^{\infty}$ be an infinite sequence of planar graphs indexed by treewidth. Let $f = \{f_k\}_{k=1}^{\infty}$ be any (possibly non-uniform) sequence of algorithms that perform certain tasks as specified below. Then we have the following results.*

(a) Suppose f solves a decision version of the MAP estimation problem in binary graphical models defined on $\{\mathcal{G}_k\}$, i.e., deciding whether $\log \psi^{\mathrm{MAP}} \geq c$ for any given c. Assuming that **NP** $\not\subseteq$ **P/poly**, *we have that $\beta_f(\mathcal{G}_k, 2)$ is super-polynomial with respect to k. Further, assuming the exponential-time hypothesis we have that $\beta_f(\mathcal{G}_k, 2) = 2^{\Omega(k^{1/2})}$.*

[6] The class **BPP/poly** refers to the class of problems which are solvable by a randomized polynomial time non-uniform algorithm. It is well-know in complexity theory, that randomness is not required if the underlying algorithms are non-uniform or equivalently **BPP/poly** = **P/poly**.

(b) Suppose f approximately computes the log-optimal value $\log \psi^{\text{MAP}}$ *to within a multiplicative factor of* $1 \pm \varepsilon$ *in binary graphical models defined on* $\{\mathcal{G}_k\}$*. Assuming the exponential-time hypothesis, we have that* $\beta_f(\mathcal{G}_k, 2)$ *cannot be polynomial in both* k *and* ε^{-1}*, i.e.,* f *cannot be a fully-polynomial time approximation scheme.*

(c) Suppose f solves a decision version of the partition function computation problem in binary graphical models defined on $\{\mathcal{G}_k\}$*, i.e., deciding whether* $Z(\psi) \geq c$ *for any given* c*. Assuming that* $\mathbf{NP} \not\subseteq \mathbf{P/poly}$*, we have that* $\beta_f(\mathcal{G}_k, 2)$ *is super-polynomial with respect to* k*. Further, assuming the exponential-time hypothesis we have that* $\beta_f(\mathcal{G}_k, 2) = 2^{\Omega(k^{1/2})}$*.*

(d) Suppose f approximately computes the partition function $Z(\psi)$ *to within a multiplicative factor of* $1 \pm \varepsilon$ *in graphical models with ternary variables defined on* $\{\mathcal{G}_k\}$*. Assuming that* $\mathbf{NP} \not\subseteq \mathbf{P/poly}$*, we have that f cannot have runtime polynomial in the treewidth k,* ε^{-1}*,* and *in* $\log(\frac{1}{\delta})$*, i.e., f cannot be a fully-polynomial time randomized approximation scheme.*

Proof The proofs of these statements are similar to those of Theorems 5.3, 5.4, 5.5, and 5.6, but the grid-minor hypothesis is not required due to Theorem 2.2. ■

Results by Demaine et al. (2005) show that the grid-minor hypothesis actually holds for bounded-genus graphs, of which planar graphs are a special case (planar graphs have genus 0). Therefore, Theorem 5.7 holds more generally for the inference problem in families of graphical models defined on bounded-genus graphs that have unbounded treewidth.

6 IMPLICATIONS

Our results show that inference is intractable in any family of graphical models with unbounded treewidth. An interesting point to note here is that we measure complexity with respect to the *treewidth* parameter rather than with respect to graph size. The following example will make this distinction clear: suppose that one considers a family of graphs $\{\mathcal{G}_k\}$ indexed by treewidth k such that the size of each graph $\mathcal{G}_k$ is exponential with respect to the treewidth k. In such a family of graphical models, the junction tree algorithm provides an inference procedure that is *polynomial* with respect to graph size (see also (Becker and Geiger, 2001) for more details). However this conclusion does not contradict our analysis, as the runtime is still *super-polynomial* with respect to treewidth. More generally, suppose that there exists a graph parameter $\alpha(\mathcal{G})$ for a graph $\mathcal{G}$ such that inference is tractable with respect to α in some family of models defined on graphs $\{\mathcal{G}_k\}$ with unbounded treewidth – that is, the runtime of inference is polynomial with respect to α. Our results allow us to conclude that α *must be super-polynomial* with respect to treewidth for graphs in the family $\{\mathcal{G}_k\}$. Otherwise, one would have a family of graphical models with unbounded treewidth in which inference is tractable with respect to treewidth.

Our results state that treewidth is the *key structural parameter* that dictates the complexity of inference – inference is easy in *all* families of models with bounded treewidth, and it is intractable in *every* family of models with unbounded treewidth. Therefore, one can relate treewidth to the other graph parameters and deduce the complexity of inference with respect to these parameters. For example, another notion of "width" considered by graph theorists is branchwidth (Robertson and Seymour, 1991), and one can show that the branchwidth of a graph is linearly bounded above and below by the treewidth of a graph; consequently inference is tractable only in bounded branchwidth models and is intractable in all families of models with unbounded branchwidth.

7 CONCLUSION

Graphical models have been widely used in numerous applications throughout machine learning, as well as in image processing, statistical physics, and networking. In many of these areas *inference* problems such as MAP estimation and marginal/partition function computation are commonly encountered. As a result several exact and approximate inference procedures have been proposed, and it is important to understand conditions under which these procedures are tractable. In this paper we studied the question of whether there is any structural property of a graphical model, other than treewidth, that enables tractable inference. In other words, we investigated whether there is some class of models with unbounded treewidth in which inference is tractable. Using various concepts and recent results from complexity theory and graph theory, we conclude that it is unlikely that an alternate structural property exists, i.e., inference is hard in *every* family of models with unbounded treewidth.

We believe that providing analysis that is conditional on the grid-minor hypothesis of Robertson et al. (1994) gives compelling evidence for our results. Contradicting these results would imply that the grid-minor hypothesis of Robertson and Seymour is false. Nevertheless it would be of great interest to prove the results of this paper without resorting to the grid-minor hypothesis. Indeed, recent work by Reed and Wood (2008) has shown the existence of grid-like minors in all graphs with polynomial treewidth. More refined analysis than that presented in this paper could perhaps be employed to use these recent developments to prove our results without resorting to the grid-minor hypothesis. Finally, our conclusions are based purely on structural properties of the graph and assume worst-case potentials in the graphical model. Measures of "hardness of potentials" have also been proposed and studied in some contexts; one prominent example is the dynamic range of a potential function (Ihler et al., 2005). An interesting question for future investigation is the complexity of inference with respect to both structural aspects of the graphical model such as treewidth as well as some measure of hardness of the potentials.

Acknowledgments

We would like to thank Lance Fortnow and Jaikumar Radhakrishnan for helpful discussions and referring us to the result of Tamassia and Tollis (1989).

References

Ann Becker and Dan Geiger. A sufficiently fast algorithm for finding close to optimal clique trees. *Artificial Intelligence*, 125(1–2):3–17, 2001.

Béla Bollobás. *Modern Graph Theory*, volume 184 of *Graduate Texts in Mathematics*. Springer Verlag, 1998. ISBN 978-0-387-98491-9.

L. Sunil Chandran and C. R. Subramanian. A spectral lower bound for the treewidth of a graph and its consequences. *Information Processing Letters*, 87(4):195–200, 2003.

Venkat Chandrasekaran, Nathan Srebro, and Prahladh Harsha. Complexity of inference in graphical models. In *Proc. 24th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2008. URL `http://uai.sis.pitt.edu/papers/08/p70-chandrasekaran.pdf`.

Michael Chertkov and Vladimir Y Chernyak. Loop series for discrete statistical models on graphs. *Journal of Statistical Mechanics*, 2006(06):P06009, 2006.

Gregory F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2–3):393–405, 1990.

Robert G. Cowell, A. Philip Dawid, Steffen L. Lauritzen, and David J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer, 1999.

Rina Dechter and Judea Pearl. Tree clustering for constraint networks. *Artificial Intelligence*, 38 (3):353–366, 1989.

Erik D. Demaine, Fedor V. Fomin, Mohammad Taghi Hajiaghayi, and Dimitrios M. Thilikos. Subexponential parameterized algorithms on bounded-genus graphs and H-minor-free graphs. *Journal of the ACM*, 52(6):866–893, 2005.

Erik D. Demaine, MohammadTaghi Hajiaghayi, and Ken ichi Kawarabayashi. Algorithmic graph minor theory: Improved grid minor bounds and Wagner's contraction. *Algorithmica*, 54(2): 142–180, 2009.

Eugene C. Freuder. Complexity of k-tree structured constraint satisfaction problems. In *Proc. 8th National Conference on Artificial Intelligence (AAAI)*, pages 4–9, 1990. URL http://www.aaai.org/Library/AAAI/1990/aaai90-001.php.

Leonidas J. Guibas, John Hershberger, Joseph S. B. Mitchell, and Jack Snoeyink. Approximating polygons and subdivisions with minimum link paths. *International Journal of Computational Geometry and Applications*, 3(4):383–415, 1993.

Alexander T. Ihler, John W. Fisher III, and Alan S. Willsky. Loopy belief propagation: Convergence and effects of message errors. *Journal of Machine Learning Research*, 6:905–936, 2005. URL http://www.jmlr.org/papers/v6/ihler05a.html.

Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which problems have strongly exponential complexity? *Journal of Computer and System Sciences*, 63(4):512–530, 2001.

Tony Jebara. Map estimation, message passing, and perfect graphs. In *Proc. 25th Conference in Uncertainty in Artificial Intelligence (UAI)*, 2009. URL http://uai.sis.pitt.edu/papers/09/p258-jebara.pdf.

Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Lectures in Mathematics. ETH Zürich. Birkhäuser, 2003.

Richard M. Karp and Richard J. Lipton. Some connections between nonuniform and uniform complexity classes. In *Proc. 12th Annual ACM Symposium on Theory of Computing (STOC)*, pages 302–309, 1980.

Steffen L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

David Litchenstein. Planar formulae and their uses. *SIAM Journal of Computing*, 11(2):329–343, 1982.

Dániel Marx. Can you beat treewidth? In *Proc. 48th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 169–179, 2007a.

Dániel Marx. On the optimality of planar and geometric approximation schemes. In *Proc. 48th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 338–348, 2007b.

Bruce Reed and David R. Wood. Polynomial treewidth forces a large grid-like minor, 2008.

Neil Robertson and Paul D. Seymour. Graph minors. I. Excluding a forest. *Journal of Combinatorial Theory, Series B*, 35(1):39–61, 1983.

Neil Robertson and Paul D. Seymour. Graph minors. II. Algorithmic aspects of tree-width. *Journal of Algorithms*, 7(3):309–322, 1986.

Neil Robertson and Paul D. Seymour. Graph minors. X. Obstructions to tree-decomposition. *Journal of Combinatorial Theory, Series B*, 52(2):153–190, 1991.

Neil Robertson, Paul D. Seymour, and Robin Thomas. Quickly excluding a planar graph. *Journal of Combinatorial Theory, Series B*, 62(2):323–348, 1994.

Dan Roth. On the hardness of approximate reasoning. *Artificial Intelligence*, 82(1-2):273–302, 1996.

Larry Stockmeyer. Planar 3-colorability is polynomial complete. *SIGACT News*, 5(3):19–25, 1973.

Roberto Tamassia and Ioannis G. Tollis. Planar grid embedding in linear time. *IEEE Transactions on Circuits and Systems*, 36(9):1230–1234, September 1989.

Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305, 2008.